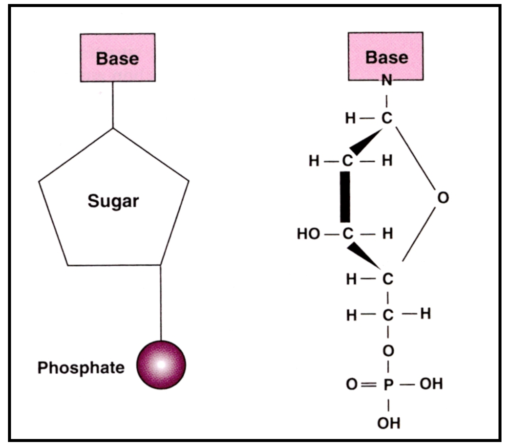
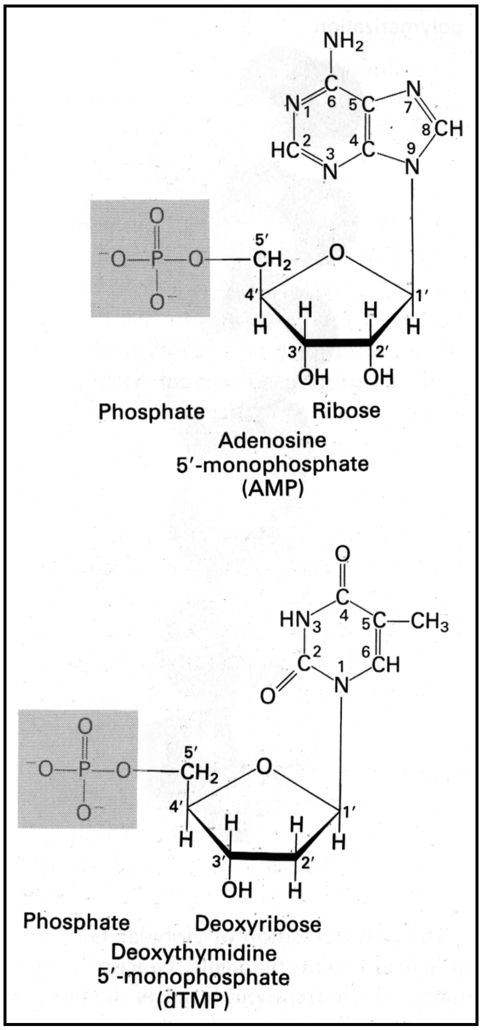
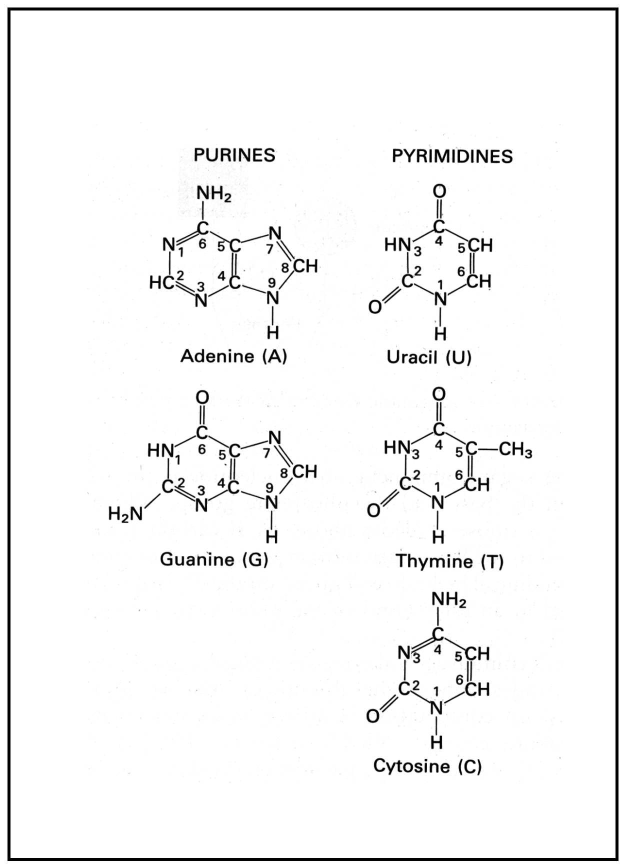
# NUCLEIC ACIDS

[Figures from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, Illinois, USA.]

Nucleic acids are comprised of chemically linked sequences of **nucleotides**. Each nucleotide is comprised of a cyclical nitrogen-containing **base**, a 5-carbon **sugar**, and a **phosphate** group. It is the sequences of the bases that constitutes the genetic information encoded in the genome.

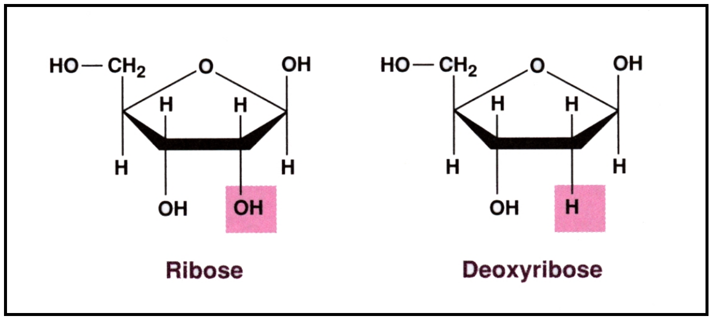


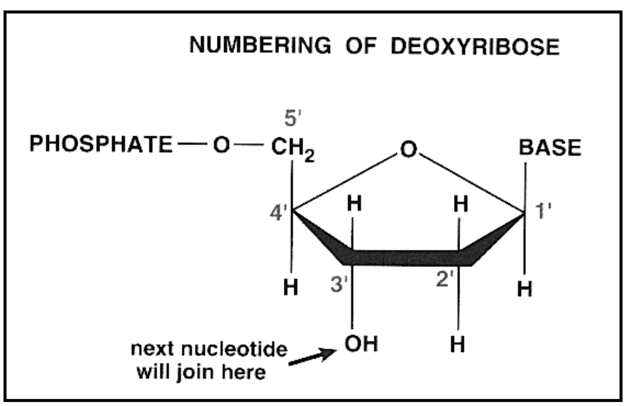

There are two types of bases - **purines** (adenine and guanine) and **pyrimidines** (cytosine and thymine in DNA and cytosine and uracil in RNA). Purines are fused 5- and 6-member rings and pyrimidines are 6-member rings. These bases constitute the genetic alphabet and consist of the letter abbreviations A, T, C, and G in DNA and A, U, C, and G in RNA.

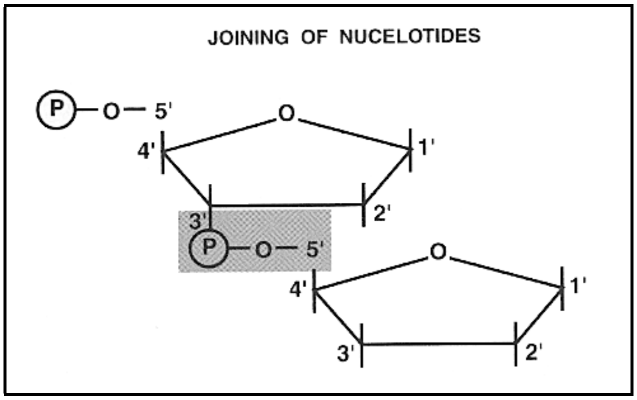
When a 5-carbon sugar (pentose) is linked to a purine or a pyrimidine, it is called a **nucleoside**. The linkage to the sugar is from the $N_1$ position of the pyrimidine or the $N_9$ position of the purine to the $C_1$ position of the sugar. Since both the nucleotides and the sugars are numbered, the positions of the carbons of the sugars are conventionally designated as prime ('). The sugar (pentose) that is present in DNA is deoxyribose and the sugar in RNA is ribose. The only difference between ribose and deoxyribose is the presence of a hydroxyl group on the 2' position of the sugar.

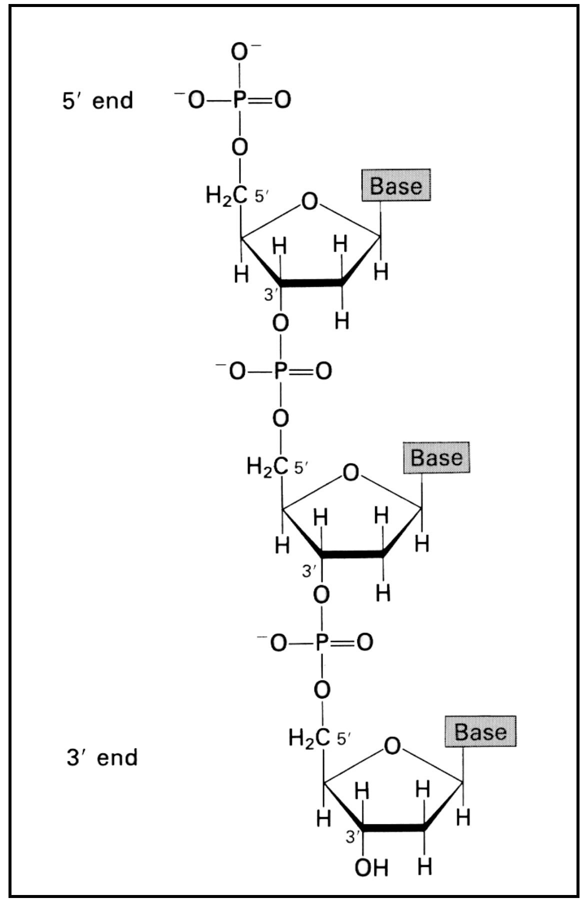



Because of the OH group of ribose, RNA is less chemically stable than DNA. RNA is thus considered base-labile.

The nucleotides (a combination of a nitrogenous base, a sugar, and a phosphate group) connect together to form a nucleic acid chain following a specific pattern of connection. The 5' position of one sugar ring is connected to the 3' position of the next sugar ring by a phosphate group. Thus, each nucleic acid chain starts with a free 5' phosphate on the sugar and ends with a 3' hydroxyl on the sugar. This 5' and 3' arrangement is reflective of the basic molecular concept of polarity (the 5' to 3' concept).



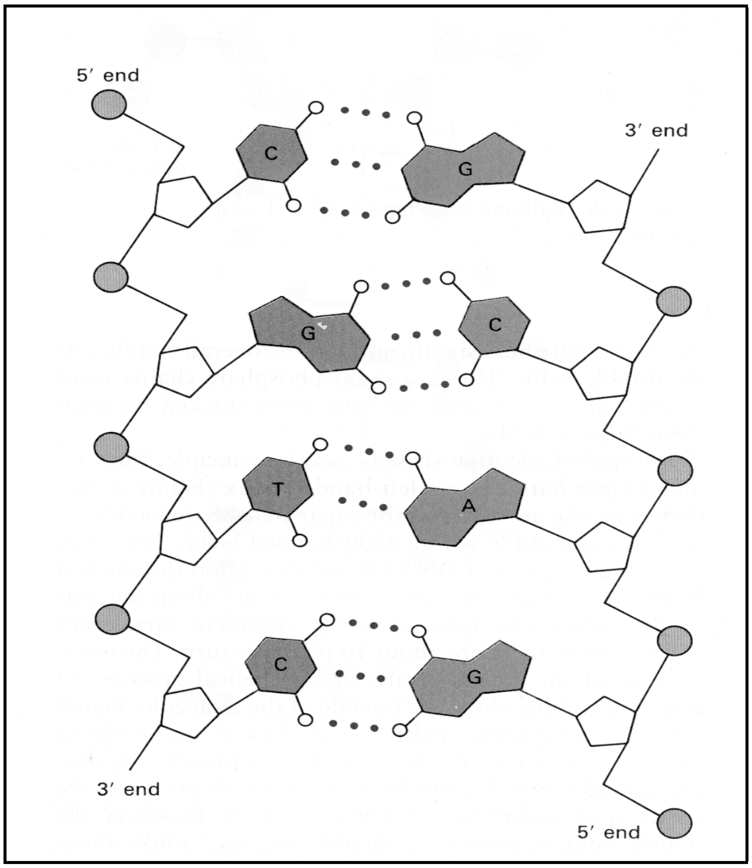



The following table sorts out the relevant nomenclature:

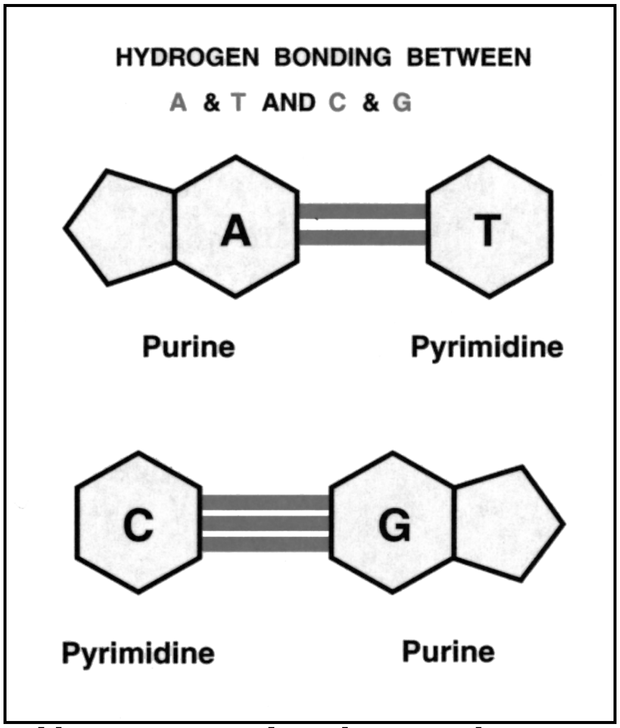
| Base | Nucleoside | Nucleotide | Abbreviations RNA | DNA |
|---|---|---|---|---|
| Purine | | | | |
| Adenine | Adenosine | Adenylic acid | AMP | dAMP |
| Guanine | Guanosine | Guanlyic acid | GMP | dGMP |
| Pyrimidine | | | | |
| Cytosine | Cytidine | Cytidylic acid | CMP | dCMP |
| Thymidine | Thymidine | Thymidylic acid | | dTMP |
| Uracil | Uridine | Uridylic acid | UMP | |

While the above nomenclature may not be of particular concern to pathologists, it is of some interest when considering two common agents that are used to label proliferating cells - **tritiated thymidine and bromodeoxyuridine**. It can be seen that these are named after the related nucleoside and, in fact, these agents are referred to as thymidine analogs and are incorporated into newly synthesized DNA in place of thymidine.

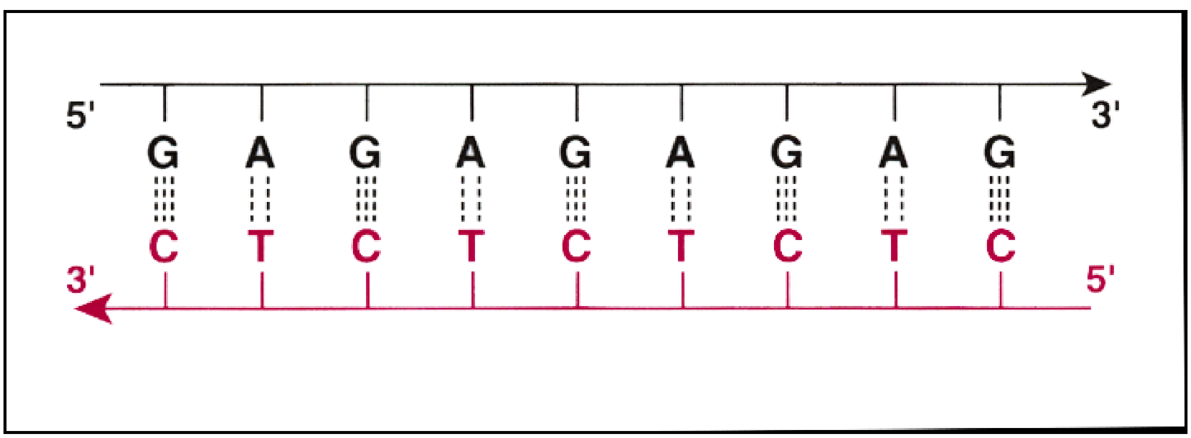



Another basic concept of molecular biology is **complementarity**. Within the double-stranded DNA helix, a purine from one strand of DNA is always paired with a pyrimidine from the other strand. The two DNA strands are connected to one another via hydrogen bonds which form between the bases.

A & G are purines while C & T are pyrimidines. In a point mutation called a **transversion** a purine is replaced by a pyrimidine or vice versa. In a **transition**, a purine is replaced by a purine or a pyrimidine is replaced by a pyrimidine.




Adenine (A) always pairs with thymidine (T) via two hydrogen bonds while guanine (G) always pairs with cytosine (C) via three hydrogen bonds. The A-T and G-C base pairs are spoken of as being complementary. As a consequence of this specific pairing in DNA, the two strands are always equidistant between the larger purine molecule and the smaller pyrimidine molecule. The fact that there are 3 hydrogen bonds between G and C and only 2 between A and T explains why it takes more energy to separate a GC pair than to separate an AT pair. Thus, GC-rich sequences in the DNA are relatively more stable than AT-rich sequences. Because of the complementarity between A and T and between C and G, there can be accurate reproduction of genetic material as each stand can serve as a template for the synthesis of the opposite strand. Because of the complementarity between the two strands of DNA, they run in opposite directions from one another and are referred to as being **antiparallel** to each other.




## DNA Methylation

The cytosine of DNA is sometimes methylated at the C5 position to form methylcytosine. In mammals, the sites of methylation are typically cytosine bases that are followed by a guanine on the same strand. This is sometimes written as

CpG. There are two noteworthy points about DNA methylation. First, methylation on a gene, especially heavy methylation, results in reduced expression of that gene. Distinct patterns of DNA methylation are found in different tissues. Second, the pattern of methylation is passed on when DNA replicates by the action of DNA methylase. This represents a heritable epigenetic change.

Genes that are constitutively expressed in all cells (housekeeping genes) typically exist in a permanently unmethylated state. The absence of the methyl groups probably provides a signal for their continued expression. Genes that show tissue-specific control have variations in DNA methylation so that only the genes appropriate for the function of that cell are unmethylated and, therefore, will be transcribed. The absence of methyl groups on DNA is believed to be associated with the ability to be transcribed but not with the act of transcription itself.

An important regulatory role of DNA methylation occurs in the phenomenon know as **genomic imprinting** in which either a maternal or paternal allele is methylated and, thus, silenced.

## DNA versus RNA

In contrast to DNA, RNA is single-stranded and the pyrimidine uracil replaces the DNA base thymidine. The chemical difference between uracil and thymidine is the absence of a methyl group at position 5 on the ring in uracil. Another difference is that the pentose ribose is present in RNA as opposed to deoxyribose in DNA (see

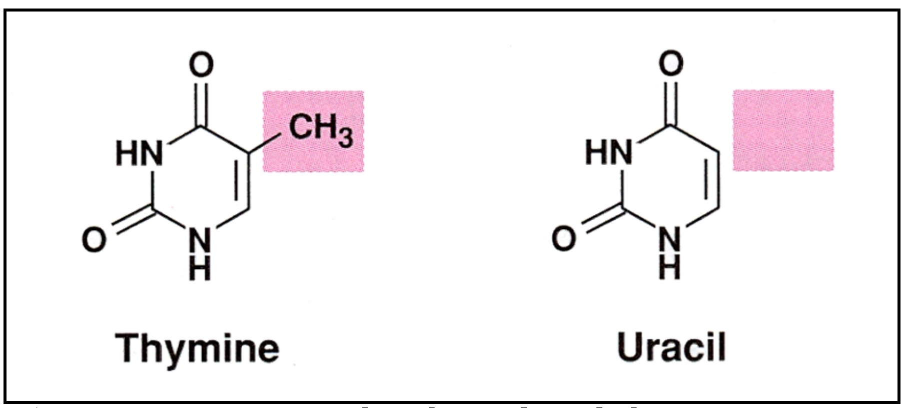



figure on 2-2). As a consequence the ribonucleotide has a 2’ OH group on the sugar while this OH group is not present in DNA. This accounts for RNA being less chemically stable than DNA. Thus, RNA is spoken of as being base labile.

## Types of RNA

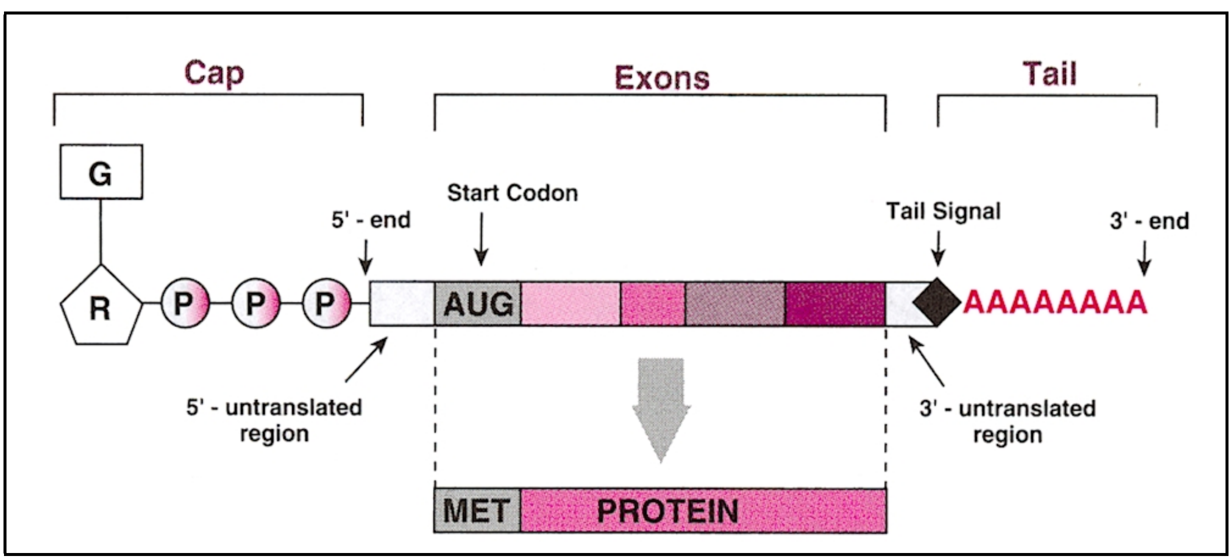
There are three types of RNA in the cell: messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). All three types are transcribed from DNA by a similar mechanism but a different enzyme (RNA polymerase) is involved with the synthesis of each type of RNA in eukaryotic cells.

Each run of three mRNA nucleotides is called a codon. Each codon encodes one amino acid. This can be thought of as letters of a 4 character alphabet spelling words that are each three letters in length. Stringing several words together makes a sentence (a string of adjacent amino acids whose sequence defines a specific protein). Thus, the genetic code is embodied in the specific codons. While some codons specify only one particular amino acid, other specific amino acids are encoded by more than one different codons. Because of this phenomenon, the genetic code is sometime referred to as degenerate (a better term might have been redundant). In addition there are three codons which are nonsense or termination codons and these function to terminate translation.

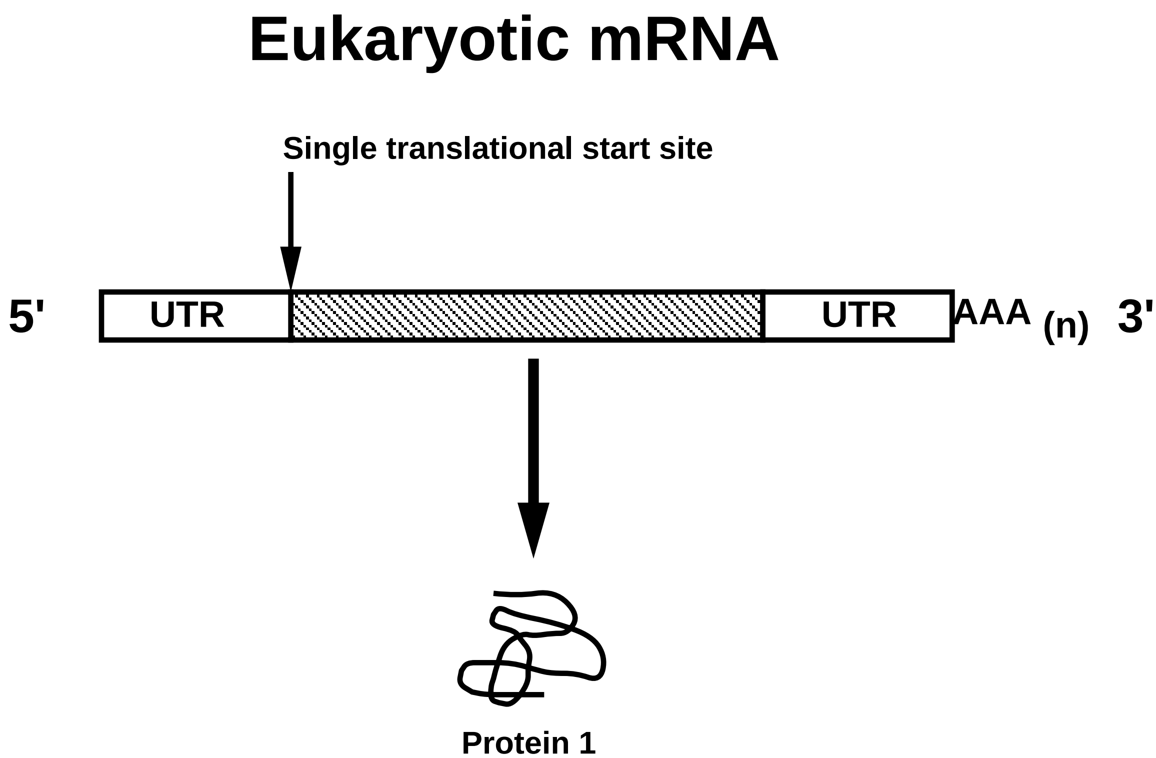
The genetic code is specified in the table below.

| | 2nd (middle) Base | | | | |
|---|---|---|---|---|---|
| 1st Base | U | C | A | G | 3rd Base |
| U | UUU Phe | UCU Phe | UAU Tyr | UGU Cys | U |
| | UUC Phe | UCC Phe | UAC Tyr | UGC Cys | C |
| | UUA Leu | UCA Leu | UAA STOP | UGA STOP | A |
| | UUG Leu | UCG Leu | UAG STOP | UGG Trp | G |
| C | CUU Leu | CCU Pro | CAU His | CGU Arg | U |
| | CUC Leu | CCC Pro | CAC His | CGC Arg | C |
| | CUA Leu | CCA Pro | CAA Gln | CGA Arg | A |
| | CUG Leu | CCG Pro | CAG Gln | CGG Arg | G |
| A | AUU Ile | ACU Thr | AAU Asn | AGU Ser | U |
| | AUC Ile | ACC Thr | AAC Asn | AGC Ser | C |
| | AUA Ile | ACA Thr | AAA Lys | AGA Arg | A |
| | AUG Mat | ACG Thr | AAG Lys | AGG Arg | G |
| G | GUU Val | GCU Ala | GAU Asp | GGU Gly | U |
| | GUC Val | GCC Ala | GAC Asp | GGC Gly | C |
| | GUA Val | GCA Ala | GAA Glu | GGA Gly | A |
| | GUG Val | GCG Ala | GAG Glu | GGG Gly | G |

mRNA molecules typically have a cap at the 5' end. This **cap** involves a methylated guanosine which is attached to the first nucleotide of the mRNA via a triphosphate group. In addition, there is a **poly(A) tail** consisting of 150 to 250 adenine nucleotides at the 3' end of the mRNA. Neither the cap nor the poly (A) tail is encoded by the DNA, but rather these are added after the start of transcription. The presence of the poly(A) tail allows for separation of mRNA from tRNA and rRNA, neither of which has a poly(A) tail.



Many mRNA molecules also have sequences at the 5' (leader) and 3' (trailer) ends which are not translated into protein and are termed **UTR**'s for untranslated regions.




Most of the RNA in a cell is **rRNA** and is located in the ribosomes. rRNA does not code for proteins but helps decode the information contained in the mRNA.

There are many types of **tRNA** in the cell. Like the rRNA, tRNA also helps to decode mRNA but is not itself translated into protein. Each of the different tRNAs recognizes and binds to one of the 20 amino acids and interacts with mRNA by virtue of its unique three dimensional structure.